JOINT SAMPLING DISTRIBUTION OF MEAN AND STANDARD DEVIATION FOR A CHI-SQUARE UNIVERSE

BY

MELVIN DALE SPRINGER
B.S., University of Illinois, 1940
M.S., University of Illinois, 1941

AN ABSTRACT OF A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN
MATHEMATICS IN THE GRADUATE COLLEGE
OF THE UNIVERSITY OF ILLINOIS, 1947

URBANA, ILLINOIS
1950

JOINT SAMPLING DISTRIBUTION OF MEAN AND STANDARD DEVIATION FOR A CHI-SQUARE UNIVERSE

BY

MELVIN DALE SPRINGER

B.S., University of Illinois, 1940
M.S., University of Illinois, 1941

AN ABSTRACT OF A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN
MATHEMATICS IN THE GRADUATE COLLEGE
OF THE UNIVERSITY OF ILLINOIS, 1947

URBANA, ILLINOIS
1950

ACKNOWLEDGMENT

The author is deeply indebted to Professor E. L. Welker for suggesting the problem of this paper, and for his inspiration and guidance in bringing it to a conclusion.

R. UNIV.
BIBLIOTHEEK
LEIDEN

JOINT SAMPLING DISTRIBUTION OF MEAN AND STANDARD DEVIATION FOR A CHI-SQUARE UNIVERSE

I. History and Introduction

R. A. Fisher has shown that the joint sampling distribution of mean and standard deviation for a normal universe is*

$$F(\bar{x},s) = Ke^{\frac{-n\{s^2+(\bar{x}-M)^2\}}{2\sigma^2}} s^{n-2}, \quad -\infty < \bar{x} < \infty, \quad -\infty < s < \infty,$$

from which it follows immediately that $\bar{x}$ and s are independent. For non-normal universes, however, the mean and standard deviation are not independent. In fact, Geary has proved that a necessary and sufficient condition that $\bar{x}$ and s be independent in the probability sense is that the x's be normally distributed in the parent population. The dependence of $\bar{x}$ and s makes the determination of their joint sampling distribution for non-normal universes considerably more difficult. A. T. Craig has found in the form of integrals the simultaneous frequency function of $\bar{x}$ and s for a general type of non-normal parent population when $n=2$, 3, or 4; while Baker has derived the joint distribution of mean and standard deviation for samples of two from a non-normal universe consisting of two normal components. Aside from this, however, very little is known concerning the joint sampling distribution of mean and standard deviation for non-normal universes. It is the purpose of this paper to determine the joint sampling distribution of mean and standard deviation for a Chi-Square universe, with special emphasis on the universe characterized by two degrees of freedom. For this latter universe and for certain special cases of the universe with four degrees of freedom, we shall determine: (1) the algebraic form of the joint distribution of $\bar{x}$ and s; (2) the effect of n upon the correlation between $\bar{x}$ and s; (3) the regression equations, scedastic functions, and indexes of correlation; (4) the more important geometric properties of the correlation surfaces.

* Throughout this discussion, $\bar{x}$ denotes the sample mean, M the mean of the universe, s the standard deviation of the sample, σ the standard deviation of the universe, and n the size of the sample.

II. Joint Sampling Distribution of Mean and Standard Deviation for a Chi-Square Universe with Two Degrees of Freedom

Consider the universe whose distribution function is

$$f(x) = \frac{1}{2^{\frac{r}{2}}\,\Gamma\left(\frac{r}{2}\right)} x^{\frac{r}{2}-1}e^{-\frac{1}{2}x},\ 0 \leq x < \infty.$$

This is the general Chi-Square universe with r degrees of freedom. In the present chapter we are concerned with the particular universe

$$f(x) = \frac{1}{2}e^{-\frac{1}{2}x},\ 0 \leq x < \infty$$

having two degrees of freedom. We inquire as to the algebraic form of the joint distribution of ${}^{n}X$ and ${}^{n}\Sigma$.*

We approach this problem by drawing a sample of two individuals, x_1 and x_2, from the infinite parent universe with two degrees of freedom. The probability that x_1 and x_2 fall simultaneously within x_1+dx_1 and x_2+dx_2, respectively, is given, to within infinitesimals of higher order, by

(I): $$f(x_1)f(x_2)dx_1dx_2 = \frac{1}{4}e^{-\frac{1}{2}(x_1+x_2)}dx_1dx_2.$$

For samples of two,

(II):
$$^{2}X = \frac{1}{2}(x_1 + x_2)$$
$$^{2}\Sigma = \frac{1}{2}(x_1 - x_2)$$

or inversely,

(III):
$$x_1 = {}^{2}X + {}^{2}\Sigma$$
$$x_2 = {}^{2}X - {}^{2}\Sigma.$$

Now, (II) may be regarded as a transformation upon the varieties x_1 and x_2, which when applied to (I) gives

$$f(x_1)f(x_2)dx_1dx_2 = K_2e^{-{}^{2}x}d\,{}^{2}\Sigma\,d\,{}^{2}X,\ 0 \leq {}^{2}X < \infty,\ 0 \leq {}^{2}\Sigma < \infty,$$

and $0 \leq {}^{2}\Sigma \leq {}^{2}X'$ for a fixed ${}^{2}X = {}^{2}X'$.

Thus, the probability that ${}^{2}X$ and ${}^{2}\Sigma$ fall simultaneously within ${}^{2}X+d\,{}^{2}X$ and ${}^{2}\Sigma+d\,{}^{2}\Sigma$, respectively, is given (to within infinitesimals of higher order) by

$$F_2({}^{2}\Sigma,{}^{2}X)d\,{}^{2}\Sigma\,d\,{}^{2}X = K_2e^{-{}^{2}x}d\,{}^{2}\Sigma\,d\,{}^{2}X,\ 0 \leq {}^{2}X < \infty,\ 0 \leq {}^{2}\Sigma < \infty,$$

$0 \leq {}^{2}\Sigma \leq {}^{2}X'$ for a fixed ${}^{2}X = {}^{2}X'$,

---

* Throughout this paper, ${}^{n}X$ and ${}^{n}\Sigma$ denote the mean and standard deviation, respectively, for a sample of n items, while $F_n({}^{n}\Sigma, {}^{n}X)$ represents the joint sampling distribution of ${}^{n}\Sigma$ and ${}^{n}X$.

where K_2 is a constant so determined that

$$\int_0^\infty \int_0^{{}^2X} K_2 e^{-{}^2X} d\,{}^2\Sigma\, d\,{}^2X = 1.$$

${}^2X$ and ${}^2\Sigma$ are, therefore, distributed jointly in accordance with

$$F_2({}^2\Sigma, {}^2X) = K_2 e^{-{}^2X},\ 0 \leq {}^2X < \infty,\ 0 \leq {}^2\Sigma < \infty,\ 0 \leq {}^2\Sigma \leq {}^2X'$$

for a fixed ${}^2X = {}^2X'$.

The form of $F_2({}^2\Sigma, {}^2X)$ might at first suggest that ${}^2\Sigma$ and ${}^2X$ are independent. A closer examination, however, reveals that such is not the case. In fact, ${}^2\Sigma$ and ${}^2X$ cannot possibly be independent, since the range of ${}^2\Sigma$ depends upon ${}^2X$. That is, for a given value of ${}^2X$, the range of ${}^2\Sigma$ is from 0 to ${}^2X$. More generally, for a sample of size n, the range of ${}^n\Sigma$ is from 0 to $\sqrt{n-1}\,{}^nX$. This is not obvious, but can be readily established by geometric methods.

We return now to the problem of determining the distribution $F({}^n\Sigma, {}^nX)$. Of fundamental importance throughout the derivation of this distribution is the transformation

$$T_n: \begin{cases} {}^n\Sigma = \dfrac{\sqrt{n-1}}{n} \sqrt{n^{n-1}\Sigma^2 + ({}^{n-1}X - X_n)^2} \\ {}^nX = \dfrac{(n-1)^{n-1}X + X_n}{n} \\ X_n = X_n \end{cases}$$

which is a consequence of the definitions of ${}^nX$ and ${}^n\Sigma$. Or inversely,

$$\begin{cases} {}^{n-1}X = \dfrac{n\,{}^nX - X_n}{n-1} \\ {}^{n-1}\Sigma^2 = \dfrac{n}{(n-1)^2}\left[(n-1)^n\Sigma^2 - ({}^nX - X_n)^2\right] \end{cases}$$

which lead to the inverse of T_n, namely

$$T_n^{-1}: \begin{cases} {}^{n-1}\Sigma = \dfrac{\sqrt{n}}{n-1} \sqrt{(n-1)^n\Sigma^2 - ({}^nX - X_n)^2} \\ {}^{n-1}X = \dfrac{1}{n-1}(n\,{}^nX - X_n) \\ X_n = X_n. \end{cases}$$

To obtain $F_3({}^3\Sigma,{}^3X)$, we multiply $F_2({}^2\Sigma,{}^2X)d\,{}^2\Sigma\,d\,{}^2X$ by $f(x_3)dx_3$, apply the transformation T_3, and integrate out the variable x_3. That is:

$$F_2({}^2\Sigma,{}^2X)d\,{}^2\Sigma\,d\,{}^2Xf(X_3)dX_3 = K_2e^{-2x}\cdot\frac{1}{2}\,e^{-\frac{1}{2}x_3}d\,{}^2\Sigma\,d\,{}^2XdX_3$$

$$= \frac{K_2}{2}\,e^{-\frac{3}{2}{}^3X}\left|J_3\right|d\,{}^3\Sigma\,d\,{}^3XdX_3{}^*$$

$$= G_2({}^3\Sigma,{}^3X,X_3)d\,{}^3\Sigma\,d\,{}^3XdX_3.$$

$$F_3({}^3\Sigma,{}^3X)d\,{}^3\Sigma\,d\,{}^3X = \int_{{}^3X-\sqrt{2}\,{}^3\Sigma}^{{}^3X+\sqrt{2}\,{}^3\Sigma}\left[G({}^3\Sigma,{}^3X,X_3)d\,{}^3\Sigma\,d\,{}^3X\right]dX_3$$

$$= K_3{}^3\Sigma e^{-\frac{3}{2}{}^3X}d\,{}^3\Sigma\,d\,{}^3X.$$

Thus, the probability that ${}^3X$ and ${}^3\Sigma$ fall simultaneously within the intervals (a,b) and (c,d), respectively, is given by

$$F({}^3\Sigma,{}^3X)d\,{}^3\Sigma\,d\,{}^3X = K_3{}^3\Sigma e^{-\frac{3}{2}{}^3X}d\,{}^3\Sigma\,d\,{}^3X,$$

where again the constant K_3 is so determined that

$$\int_0^\infty\int_0^{\sqrt{2}\,{}^3X} K_3{}^3\Sigma e^{-\frac{3}{2}{}^3X}d\,{}^3\Sigma\,d\,{}^3X = 1.$$

${}^3X$ and ${}^3\Sigma$ are, therefore, distributed jointly in accordance with

$$F_3({}^3\Sigma,{}^3X) = K_3{}^3\Sigma e^{-\frac{3}{2}{}^3X},\ 0 \le {}^3X < \infty,\ 0 \le {}^3\Sigma < \infty,$$

$$0 \le {}^3\Sigma \le \sqrt{2}\,{}^3X' \text{ for a fixed } {}^3X = {}^3X'.$$

Similarly, by multiplying $F_3({}^3\Sigma,{}^3X)d\,{}^3\Sigma\,d\,{}^3X$ by $f(x_4)dx_4$, we find that the joint distribution of ${}^4X$ and ${}^4\Sigma$ is

$$F_4({}^4\Sigma,{}^4X) = K_4{}^4\Sigma^2e^{-2\,{}^4X},\ 0 \le {}^4X < \infty,\ 0 \le {}^4\Sigma < \infty,$$

$$0 \le {}^4\Sigma \le \sqrt{3}\,{}^4X' \text{ for a fixed } {}^4X = {}^4X'.$$

Note that $F_2({}^2\Sigma,{}^2X)$, $F_3({}^3\Sigma,{}^3X)$, and $F_4({}^4\Sigma,{}^4X)$ are all of the form

$$F_n({}^n\Sigma,{}^nX) = K_n{}^n\Sigma^{n-2}e^{-\frac{n}{2}},\ 0 \le {}^nX < \infty,\ 0 \le {}^n\Sigma < \infty,$$

$$0 \le {}^n\Sigma \le \sqrt{n-1}\,{}^nX' \text{ for a fixed } {}^nX = {}^nX',\ n = 2, 3, 4$$

---

* Throughout this paper J_n, $n=1, 2, \ldots$, denotes the Jacobian of the transformation T_n^{-1}.

where K_n is a positive constant determined in each case so that

$$\int_0^{\infty}\int_0^{\sqrt{n-1}\,{}^nX} K_n\,{}^n\Sigma^{n-2}e^{-\frac{n}{2}\,{}^nX}d\,{}^n\Sigma\,d\,{}^nX = 1.*$$

It would seem, therefore, that $F_n({}^n\Sigma,{}^nX)$, $n>1$, all obey the general law

(IV): $F({}^n\Sigma,{}^nX) = K_n\,{}^n\Sigma^{n-2}e^{-\frac{n}{2}\,{}^nX}$, $n > 1$, $0 \leq {}^nX < \infty$,

$0 \leq {}^n\Sigma < \infty$, $0 \leq {}^n\Sigma \leq \sqrt{n-1}\,{}^nX'$ for a fixed ${}^nX = {}^nX'$.†

The remainder of this chapter is devoted to establishing this conjecture by mathematical induction.

Having found the joint sampling distribution of ${}^nX$ and ${}^n\Sigma$ to be of the form

$$F({}^n\Sigma,{}^nX) = K_n\,{}^n\Sigma^{n-2}e^{-\frac{n}{2}\,{}^nX},\ 0 \leq {}^nX < \infty,\ 0 \leq {}^n\Sigma < \infty,$$

$$0 \leq {}^n\Sigma \leq \sqrt{n-1}\,{}^nX' \text{ for a fixed } {}^nX = {}^nX'$$

one can easily show that K_n has the value

$$K_n = \left(\frac{n}{2}\right)^n \frac{1}{(n-1)^{\frac{n-3}{2}}\,\Gamma(n)}.$$

Also, the marginal distributions of ${}^nX$ and ${}^n\Sigma$ are, respectively,

$$g({}^nX) = \frac{\left(\frac{n}{2}\right)^n {}^nX^{n-1}\,e^{-\frac{n}{2}\,{}^nX}}{\Gamma(n)}$$

---

* The constant K_n in the distributions $F({}^n\Sigma,\ {}^nX) = K_n\,{}^n\Sigma^{n-2}e^{-\frac{n}{2}\,{}^nX}$, $n=2, 3, \ldots$, is quite arbitrary, inasmuch as changing such a constant factor does not essentially alter the distribution of $F({}^n\Sigma,\ {}^nX)$. It is convenient, however, to choose K_n such that $\int_0^{\infty}\int_0^{\sqrt{n-1}\,{}^nX} K_n\,{}^n\Sigma^{n-2}\,e^{\frac{n}{2}\,{}^nX}d\,{}^n\Sigma\,d\,{}^nX = 1$, for then $F({}^n\Sigma,\,{}^nX)d\,{}^n\Sigma\,d\,{}^nX$ gives the probability, to within infinitesimals of higher order, that ${}^nX$ and ${}^n\Sigma$ occur simultaneously within the intervals ${}^nX+d\,{}^nX$ and ${}^n\Sigma+d\,{}^n\Sigma$, respectively. Thus, whenever K_i $(i=1, 2, \ldots, m, m+1, \ldots, n, \ldots)$ appears throughout the remainder of this discussion, it shall be construed to mean a constant such that $\int_0^{\infty}\int_0^{\sqrt{n-1}\,{}^nX} K_i\,{}^i\Sigma^{i-2}e^{-\frac{i}{2}\,{}^iX}d\,{}^i\Sigma d\,{}^iX = 1.$

† In future reference to similar functions, the range of one variable for a fixed value of the other will be understood but not explicitly stated.

and

$$h({}^n\Sigma) = \frac{\left(\frac{n}{2}\right)^{n-1} {}^n\Sigma^{n-2}\, e^{-\frac{n\,{}^n\Sigma}{2\sqrt{n-1}}}}{(n-1)^{\frac{n-3}{2}}\,\Gamma(n)}.$$

Finally, since $d\,{}^n\Sigma^2 = 2\,{}^n\Sigma d\,{}^n\Sigma$, it follows that the joint distribution of mean and variance is given by

$$G({}^n\Sigma^2, {}^nX) = \left(\frac{1}{2}\right)\left(\frac{n}{2}\right)^n \frac{({}^n\Sigma^2)^{\frac{n-3}{2}}\, e^{-\frac{n}{2}\,{}^nX}}{(n-1)^{\frac{n-3}{2}}\,\Gamma(n)},\quad 0 \leq {}^nX < \infty,\ 0 \leq {}^n\Sigma < \infty,$$

$0 \leq {}^n\Sigma^2 \leq (n-1)\,{}^nX'^2$ for a fixed ${}^nX = {}^nX'$.

III. Correlation Properties of the Joint Distribution

$$F({}^n\Sigma, {}^nX) = \frac{\left(\frac{n}{2}\right)^n {}^n\Sigma^{n-2}\, e^{-\frac{n}{2}\,{}^nX}}{(n-1)^{\frac{n-3}{2}}\,\Gamma(n)}$$

In discussing the correlation properties of the joint distribution, we shall have occasion to use moments about the origin of ${}^nX$ and ${}^n\Sigma$, as well as moments about the means ${}^n\overline{X}$ and ${}^n\overline{\Sigma}$. We shall denote the general moments about the origin and means, respectively, by γ_{ij} and μ_{ij}, where

$$\gamma_{ij} = \int_0^\infty \int_0^{\sqrt{n-1}\,{}^nX} F({}^n\Sigma, {}^nX)\, {}^nX^i\, {}^n\Sigma^j d\,{}^n\Sigma d\,{}^nX$$

and

$$\mu_{ij} = \int_0^\infty \int_0^{\sqrt{n-1}\,{}^nX} ({}^nX - {}^n\overline{X})^i\, ({}^n\Sigma - {}^n\overline{\Sigma})^j\, F({}^n\Sigma, {}^nX) d\,{}^n\Sigma d\,{}^nX.$$

Since the probability density of an ${}^nX$ array of ${}^n\Sigma$'s is $\dfrac{F({}^n\Sigma, {}^nX)}{g({}^nX)}$, the mean of an ${}^nX$ array of ${}^n\Sigma$'s is

$${}^n\overline{\Sigma}_{{}^nX} = \int_0^{\sqrt{n-1}\,{}^nX} \frac{{}^n\Sigma F({}^n\Sigma, {}^nX)}{g({}^nX)}\, d\,{}^n\Sigma;$$

it is a function of ${}^nX$ and is called the regression function of ${}^n\Sigma$ on ${}^nX$. Similarly, the regression function of ${}^nX$ on ${}^n\Sigma$ is

$${}^n\overline{X}_{{}^n\Sigma} = \int_{\frac{{}^n\Sigma}{\sqrt{n-1}}}^\infty \frac{{}^nX F({}^n\Sigma, {}^nX)}{h({}^n\Sigma)}\, d\,{}^nX.$$

After carrying out the integrations, we find the regression equations to be

$$ {}^{n}\overline{\Sigma}_{{}^{n}X} = \frac{(n-1)^{\frac{3}{2}}}{n}\,{}^{n}X $$

and

$$ {}^{n}\overline{X}_{{}^{n}\Sigma} = \frac{{}^{n}\Sigma}{\sqrt{n-1}} + \frac{2}{n}. $$

Thus, both regression systems are linear.

The extent of the correlation between mean and standard deviation is indicated by the correlation coefficient $r_{{}^{n}X\,{}^{n}\Sigma}$, where

$$ r_{{}^{n}X{}^{n}\Sigma} = \frac{\displaystyle\int_0^{\infty}\int_0^{\sqrt{n-1}\,{}^{n}X} ({}^{n}X - {}^{n}\overline{X})\,({}^{n}\Sigma - {}^{n}\overline{\Sigma})\,F({}^{n}\Sigma, {}^{n}X)\,d\,{}^{n}\Sigma\,d\,{}^{n}X}{\left[\left(\displaystyle\int_0^{\infty}\int_0^{\sqrt{n-1}\,{}^{n}X} ({}^{n}\Sigma - {}^{n}\overline{\Sigma})^2\,F({}^{n}\Sigma, {}^{n}X)\,d\,{}^{n}\Sigma\,d\,{}^{n}X\right)\left(\displaystyle\int_0^{\infty}\int_0^{\sqrt{n-1}\,{}^{n}X} ({}^{n}X - {}^{n}\overline{X})^2\,F({}^{n}\Sigma, {}^{n}X)\,d\,{}^{n}\Sigma\,d\,{}^{n}X\right)\right]^{\frac{1}{2}}} $$

$$ = \frac{\mu_{11}}{\left[\mu_{o2}\mu_{2o}\right]^{\frac{1}{2}}} $$

$$ = \left(1 - \frac{1}{n}\right)^{\frac{1}{2}}. $$

Thus,

$$ \lim_{n\to\infty} r_{{}^{n}X{}^{n}\Sigma} = 1, $$

so that the mean and standard deviation approach functional dependence as $n \to \infty$.

Similarly, one can measure the correlation between mean and variance through the use of the correlation coefficient $r_{{}^{n}X{}^{n}\Sigma^2}$, which is of the form

$$ r_{{}^{n}X{}^{n}\Sigma^2} = \left[\frac{2(n-1)}{2n+1}\right]^{\frac{1}{2}}. $$

Since

$$ \lim_{n\to\infty} r_{{}^{n}X{}^{n}\Sigma^2} = 1, $$

it follows that the mean and variance approach functional dependence as the number of individuals in the sample increases indefinitely.

For the regression system of ${}^n\Sigma$ on ${}^nX$, the equation of the scedastic curve is

$$ {}^n\Sigma = \frac{S({}^nX)}{[\mu_{o2}]^{\frac{1}{2}}}, $$

where

$$ S^2({}^nX) = \int_0^{\sqrt{n-1}\,{}^nX} \frac{({}^n\Sigma - {}^n\overline{\Sigma}_{{}^nX})^2 F({}^n\Sigma, {}^nX)}{g({}^nX)}\, d\,{}^n\Sigma. $$

This reduces to

$$ {}^n\Sigma = \frac{{}^nX}{2\sqrt{n+1}}. $$

In like manner, we find the equation of the scedastic curve for the regression system of ${}^nX$ on ${}^n\Sigma$ to be

$$ {}^nX = n^{-\frac{1}{2}}. $$

Thus, one regression system is homoscedastic, while the other is heteroscedastic.

Geometrically,

$$ F({}^n\Sigma, {}^nX) = \frac{\left(\frac{n}{2}\right)^n {}^n\Sigma^{n-2} e^{-\frac{n}{2}\,{}^nX}}{(n-1)^{\frac{n-3}{2}}\,\Gamma(n)}, \quad 0 \leq {}^nX < \infty, \quad 0 \leq {}^n\Sigma < \infty, $$

$$ 0 \leq {}^n\Sigma \leq \sqrt{n-1}\,{}^nX' \text{ for a fixed } {}^nX = {}^nX', $$

represents a three dimensional surface lying wholly in the first octant. We shall refer to it as a correlation surface, inasmuch as it may be regarded as a correlation table idealized into a surface. It is bounded by four planes:

$$ Z = 0 $$

$$ Z = \frac{\left(\frac{n}{2}\right)^2 (n-2)^{n-2}\,(n-1)^{\frac{1}{2}}\, e^{-(n-2)}}{\Gamma(n)} $$

$$ {}^n\Sigma = 0 $$

$$ {}^n\Sigma = \sqrt{n-1}\,{}^nX. $$

The maximum point of the surface is the point P: $\left[2\left(1-\frac{2}{n}\right), 2\sqrt{n-1}\left(1-\frac{2}{n}\right), \frac{\left(\frac{n}{2}\right)^2 (n-1)^{\frac{1}{2}}\, e^{-(n-2)}}{\Gamma(n)}\right]$. If $n=2$, the plane ${}^2X = K$ (K = a constant) intersects the correlation surface in a straight line, while the plane ${}^2\Sigma = K$ intersects it in that arc of the exponential curve

$F(^2X) = e^{-\frac{^2X}{2}}$ included between $^2X = {}^2\Sigma = K$ and $^2X = \infty$. If $n = 3$, the intersection of the plane $^3X = K_1$ (K_1 = a constant) with the surface is the straight line

$$Z = K\,{}^3\Sigma \left[K = \frac{\left(\frac{3}{2}\right)^3 e^{-\frac{3}{2}K_1}}{\Gamma(3)} \right];$$

whereas if $n > 3$, the intersection is that arc of the $(n-2)nd$ degree parabola

$$Z = K\,{}^n\Sigma^{n-2} \left[K = \frac{\left(\frac{n}{2}\right)^n e^{-\frac{n}{2}K_1}}{(n-1)^{\frac{n-3}{2}}\,\Gamma(n)} \right]$$

included between ${}^n\Sigma = 0$ and ${}^n\Sigma = \sqrt{n-1}\,{}^nX$. Finally, if $n \geqslant 3$, the plane $^3\Sigma = K_2 > 0$ intersects the correlation surface in that arc of the exponential curve

$$Z = Ke^{-\frac{n}{2}\,{}^nX} \left[K = \frac{K_2{}^{n-2}\left(\frac{n}{2}\right)^n}{(n-1)^{\frac{n-3}{2}}\,\Gamma(n)} \right]$$

included between $^nX = \dfrac{^n\Sigma}{\sqrt{n-1}} = \dfrac{K_2}{\sqrt{n-1}}$ and $^nX = \infty$.

IV. Joint Distribution of Mean and Standard Deviation for Samples of Two from a Chi-Square Universe Four Degrees of Freedom

Using the same method as in Chapter II, we find that the joint distribution of mean and standard deviation for samples of two from a Chi-Square universe with four degree of freedom is

$$(V): \quad F_2(^2\Sigma, {}^2X) = \frac{1}{4}\,(^2X^2 - {}^2\Sigma^2)\,e^{-^2X},\ 0 \leq {}^2X < \infty,\ 0 \leq {}^2\Sigma < \infty,$$

$$0 \leq {}^2\Sigma \leq {}^2X' \text{ for a fixed } {}^2X = {}^2X'.$$

The equations of the regression curves are

$$^2\overline{\Sigma}_{^2X} = \frac{3}{8}\,{}^2X$$

and

$$^2\overline{X}_{^2\Sigma} = \frac{^2\Sigma^2 + 3\,{}^2\Sigma + 3}{^2\Sigma + 1}.$$

The correlation indexes have the values $\rho_{^2\Sigma\,^2X} = 0.567$ and $\rho_{^2X\,^2\Sigma} = 0.486$, correct to three decimal places. Both regression systems are heteroscedastic, the equations of the scedastic curves being

$$^2\Sigma = \frac{1}{4}\sqrt{\frac{19}{35}}\,{}^2X$$

and

$$^2X = \frac{1}{2(^2\Sigma + 1)}\left[\frac{2\,{}^2\Sigma^5 + 2\,{}^2\Sigma^3 + 8\,{}^2\Sigma^2 + 9\,{}^2\Sigma + 3}{^2\Sigma + 1}\right]^{\frac{1}{2}}.$$

V. Joint Distribution of Mean and Standard Deviation for Samples of Three from a Chi-Square Universe with Four Degrees of Freedom

Consider again a Chi-Square universe with four degrees of freedom; its distribution function is

$$f(x) = \frac{1}{2^2\Gamma(2)}\,xe^{-\frac{1}{2}x},\ 0 \leq x < \infty.$$

If we multiply (V) by $f(x_3)\ d\,{}^2\Sigma\ d\,{}^2X\ dx_3$, apply the transformation T_3, and integrate out the variable x_3, we obtain the joint distribution of ^3X and $^3\Sigma$; namely,

$$F_3(^3\Sigma, {}^3X) = \left(\frac{3}{2}\right)^6\left(\frac{1}{40}\right){}^3\Sigma\,{}^3X(2\,{}^3X^2 - 3\,{}^3\Sigma^2)\,e^{-\frac{3}{2}\,{}^3X},\ 0 \leq {}^3X < \infty,$$

$$0 \leq {}^3\Sigma < \infty,\ 0 \leq {}^3\Sigma \leq \frac{\sqrt{6}}{3}\,{}^3X' \text{ for a fixed } {}^3X = {}^3X'.$$

The equations of the regression curves for this joint distribution are

$$^3\overline{\Sigma}_{^3X} = \frac{8\sqrt{6}}{45}\,{}^3X$$

and

$$^3\overline{X}_{^3\Sigma} = \frac{1}{6}\,\frac{[27\sqrt{6}\,{}^3\Sigma^3 + 180\,{}^3\Sigma^2 + 96\sqrt{6}\,{}^3\Sigma + 128]}{[9\,{}^3\Sigma^2 + 6\sqrt{6}\,{}^3\Sigma + 8]}.$$

The index of correlation $\rho_{^3\Sigma^3X}$ has the value $8/\sqrt{141}$ (0.675, approximately). Both regression systems are heteroscedastic, the equations of the scedastic curves being

$$^3\Sigma = 3\sqrt{\frac{19}{376}}\,{}^3X$$

$$^3X = \frac{3}{16\,(9\,{}^3\Sigma^2 + 6\sqrt{6}\,{}^3\Sigma + 8)^{\frac{3}{2}}}\,[243{,}000\sqrt{6}\,{}^3\Sigma^7 + 186{,}624\,{}^3\Sigma^6 + 497{,}664\sqrt{6}\,{}^3\Sigma^5 + 7{,}589{,}376\,{}^3\Sigma^4 + 1{,}824{,}768\sqrt{6}\,{}^3\Sigma^3 + 3{,}465{,}243\,{}^3\Sigma^2 + 589{,}824\sqrt{6}\,{}^3\Sigma + 262{,}144]^{\frac{1}{3}}.$$

VI. Summary

In this paper, the joint sampling distribution of mean and standard deviation for a Chi-Square universe with two degrees of freedom was derived. The regression functions of the joint distribution were both found to be linear. One regression system was homoscedastic; the other, heteroscedastic. The correlation coefficient $r_{^nX^n\Sigma}$ was a function of n and approached 1 as $n \to \infty$. Likewise, the correlation between mean and variance approached 1 as $n \to \infty$. The joint distributions of means and standard deviation when $n=2$ and $n=3$ were also derived for the universe with four degrees of freedom. The correlation properties of these two distributions were then discussed.

VII. Bibliography

(1) "Student," "The Probable Error of a Mean," *Biometrika,* Vol. 6 (1908), pp. 1-25.

(2) *Ibid.*

(3) R. A. Fisher, "Applications of 'Student's' Distribution," *Metron,* Vol. 5 (1928), p. 93.

(4) J. Neyman, "On the Correlation of the Mean and Variance in Samples Drawn From an Infinite Population," *Biometrika,* Vol. 18 (1926), pp. 401-13.

(5) R. C. Geary, "The Distribution of 'Student's' Ratio for Non-Normal Samples," *Journal of the Royal Statistical Society Supplement,* Vol. 3 (1936), pp. 178-84.

(6) G. A. Baker, "The Relation Between the Means and Variances, Means Squared, and Variances in Samples from Combinations of Normal Populations," *Annals of Math. Stat.,* Vol. 2 (1931), pp. 333-54.

(7) A. T. Craig, "The Simultaneous Distribution of Mean and Standard Deviation in Small Samples," *Annals of Math. Stat.,* Vol. 3 (1932), pp. 126-40.

(8) Geary, *op. cit.,* p. 179.

(9) P. R. Rider, "On the Distribution of the Ratio of Mean to Standard Deviation in Small Samples From Non-Normal Universes," *Biometrika,* Vol. 21 (1929), pp. 124-43.

(10) V. Perlo, "On the Distribution of 'Student's' Ratio for Samples of Three Drawn From a Rectangular Distribution," *Biometrika,* Vol. 5 (1933), p. 203.

(11) Geary, *op. cit.*

(12) Geary, *op. cit.*

(13) Jack Laderman, "The Distribution of 'Student's' Ratio For Samples of Two Items Drawn From Non-Normal Universes," *Annals of Math. Stat.,* Vol. 10 (1939), pp. 376-79.

(14) W. A. Shewhart and F. W. Winters, "Small Samples — New Experimental Results," *Journal of the American Statistical Association,* Vol. 23 (1928), pp. 144-53.

(15) "Sophister," "Discussion of Small Samples Drawn From an Infinite Skew Population," *Biometrika,* Vol. 20A (1928), pp. 408-22.

(16) Karl Pearson, "Notes on Skew Frequency Surfaces," *Biometrika,* Vol. 15 (1923), pp. 222-44.

(17) *Ibid.*

(18) Seimatsu Narumi, "On the General Forms of Bivariate Frequency Distributions Which Are Mathematically Possible When Regression and Variation Are Subjected to Limiting Conditions," *Biometrika,* Vol. 15 (1923), Part I, pp. 77-88; Part II, pp. 209-21.

(19) John F. Kenney, *Mathematics of Statistics,* Vol. II, Van Nostrand & Company, New York, 1939.

(20) M. G. Kendall, *The Advanced Theory of Statistics,* Vol. I, Griffin & Company, London, 1945.

VITA

Melvin Dale Springer was born at Saybrook, Illinois, on September 12, 1918. He was graduated from Bellflower High School in 1935. In September, 1936, he entered the University of Illinois, where he received his Bachelor of Science degree in June, 1940. He spent the next year doing graduate work at the University of Illinois, where in June, 1941, he received the degree of Master of Science with a major in mathematics. From 1941-44 he held a graduate assistantship and did graduate work in the Department of Mathematics at the University of Illinois. After spending the next two years in the Navy, he returned to the University of Illinois to continue graduate work towards the degree of Doctor of Philosophy. He is a member of Pi Mu Epsilon, Kappa Delta Pi, The American Mathematical Society, The Institute of Mathematical Statistics, and The American Association for the Advancement of Science.